JOURNAL OF BACTERIOLOGY, Sept. 1990, p. 4888–4900
0021-9193/90/094888-13$02.00/0

Vol. 172, No. 9

Genetic and Sequence Organization of the *mcrBC* Locus of *Escherichia coli* K-12

D. DILA, E. SUTHERLAND, L. MORAN, B. SLATKO, AND E. A. RALEIGH*
New England Biolabs, 32 Tozer Road, Beverly, Massachusetts 01915

Received 6 February 1990/Accepted 6 June 1990

The *mcrB* (*rglB*) locus of *Escherichia coli* K-12 mediates sequence-specific restriction of cytosine-modified DNA. Genetic and sequence analysis shows that the locus actually comprises two genes, *mcrB* and *mcrC*. We show here that in vivo, McrC modifies the specificity of McrB restriction by expanding the range of modified sequences restricted. That is, the sequences sensitive to McrB$^+$-dependent restriction can be divided into two sets: some modified sequences containing 5-methylcytosine are restricted by McrB$^+$ cells even when McrC$^-$, but most such sequences are restricted in vivo only by McrB$^+$ McrC$^+$ cells. The sequences restricted only by McrB$^+$C$^+$ include T-even bacteriophage containing 5-hydroxymethylcytosine (restriction of this phage is the RglB$^+$ phenotype), some sequences containing N^4-methylcytosine, and some sequences containing 5-methylcytosine. The sequence codes for two polypeptides of 54 (McrB) and 42 (McrC) kilodaltons, whereas in vitro translation yields four products, of ~29 and ~49 (McrB) and of ~38 and ~40 (McrC) kilodaltons. The McrB polypeptide sequence contains a potential GTP-binding motif, so this protein presumably binds the nucleotide cofactor. The deduced McrC polypeptide is somewhat basic and may bind to DNA, consistent with its genetic activity as a modulator of the specificity of McrB. At the nucleotide sequence level, the G+C content of *mcrBC* is very low for *E. coli*, suggesting that the genes may have been acquired recently during the evolution of the species.

The locus known as *mcrB* was one of the first restriction systems to be discovered (33), by virtue of its action on special variants of T-even bacteriophage that incorporate 5-hydroxymethylcytosine (^{hm5}C) into their DNA without further modification (see reference 50 for a review). This locus, formerly known as *rglB* (or $r_{2,4}$) (48), was rediscovered because of difficulties encountered in cloning the genes for site-specific modification methylases associated with type II restriction-modification systems (7, 26, 40, 49). In addition to ^{hm5}C-DNA, many but not all sequences methylated by site-specific cytosine modification methylases are restricted by the system in vivo, and the consensus recognition sequence 5′G^{m}C was proposed (49). McrB is thus a sequence-specific, modification-requiring restriction system. We show here that the *mcrB* locus described above actually comprises two genes and that both are required for restriction of most the sequences previously characterized as sensitive. Thus, we will refer to the complete system as the McrBC system.

The genes encoding the system are contained within the immigration control region of the *Escherichia coli* K-12 genome. Three restriction systems are encoded within 14 kilobases (kb) here (48). The well-studied *hsdRMS* locus (20, 31, 55) encodes the multisubunit type I system *Eco*K, which recognizes a seven-base sequence and cleaves the target when the sequence is not modified. The other two systems are the flanking loci *mcrBC*, described above, and *mrr*, which mediates site-specific restriction of adenine-modified DNA (22). The sequence organization of this region, judged by Southern blot analysis of chromosomal DNA, is highly variable in enteric bacteria (12), both in the *hsd* genes specifically and in the flanking sequences. Sequence analysis presented here is consistent with recent acquisition of the *mcrBC* genes by *E. coli*, possible accounting for some of the observed variability.

* Corresponding author.

At the molecular level, restriction systems consist of sequence-specific double-stranded endonucleases, usually accompanied by a sequence-specific modification methylase. So far, four classes of endonucleases have been described. The simplest are the type II enzymes, in which the endonuclease and protective methylase activities reside in separate enzymes. These endonucleases typically act as dimers of identical subunits and require only Mg^{2+} for activity (38). One group of type II isoschizomers, typified by *Dpn*I, recognizes a modified site (28), as McrBC appears to do. In contrast, type I and type III enzymes have separate specificity subunits that recognize the DNA site and require ATP in addition to Mg^{2+} for cleavage (6). Restriction mediated by these latter systems requires the cooperation of the specificity subunit with one (type III) or two (type I) other subunits, both in vivo and in vitro, and either is stimulated by (type III) or requires (type I) *S*-adenosylmethionine as well as ATP in vitro. There are several related families of type I enzymes, and specificity subunits belonging to different members of the same family can be substituted for each other. A recently described class requires *S*-adenosylmethionine for cleavage, but not ATP, and has been designated type IV (45).

The protein sequence deduced from the DNA sequence presented here, taken together with the genetic evidence, is consistent with the view that the McrBC system differs significantly from all of these, but in organization more closely resembles the multigene, nucleotide-dependent types I and III enzymes than the single-gene, nucleotide-independent type II enzymes.

While this paper was in preparation, reports appeared (52, 53) covering similar work. Our substantial differences of fact and interpretation with these reports are discussed.

(A preliminary report of some of this work was presented at the New England Biolabs Workshop on Biological DNA Modification, May 1988 [18].)

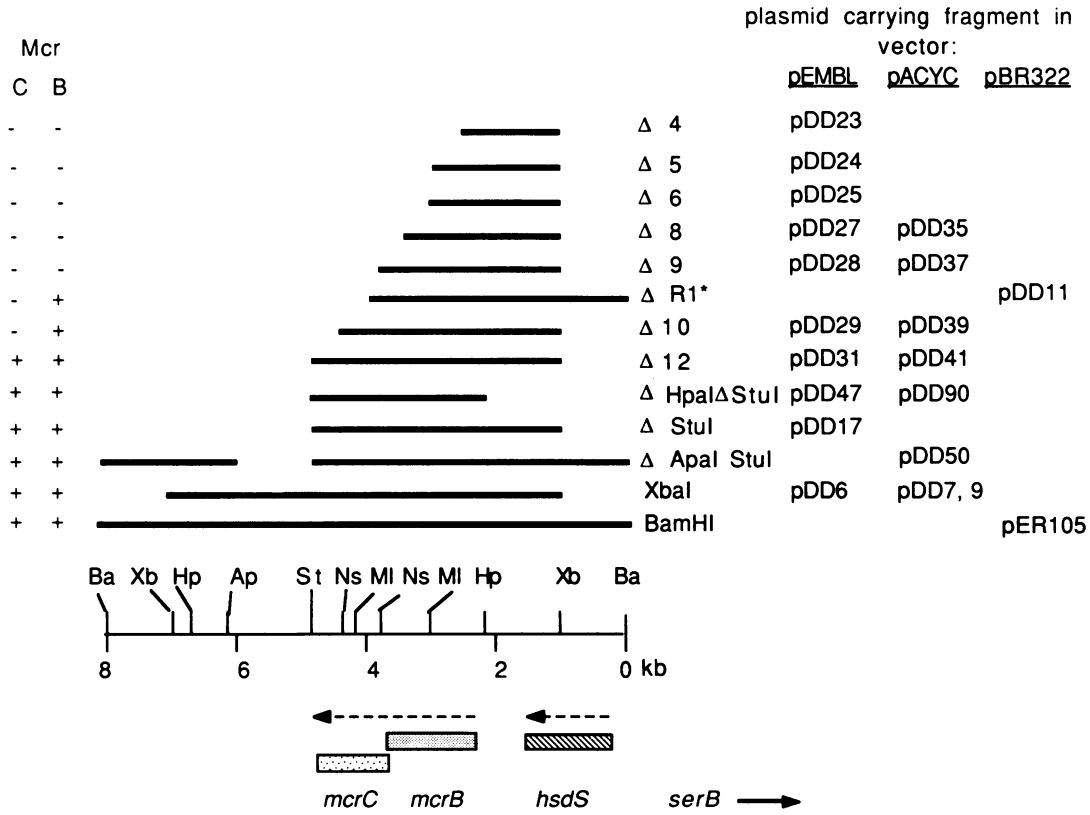

FIG. 1. Subcloning of the *mcrBC* region. A restriction map of the original cloned *Bam*HI insert is shown at the bottom, oriented as in reference 48, with the order of genes clockwise on the *E. coli* genetic map; *serB* lies an unknown distance to the right. Enzymes indicated: Ba, *Bam*HI; Xb, *Xba*I; Hp, *Hpa*I; Ap, *Apa*I; St, *Stu*I; Ns, *Nsi*I; Ml, *Mlu*I. Bars above the map indicate DNA carried in plasmids named at the right, with the vector indicated at the top of the respective column. At the left is shown the phenotype conferred on a host deleted for the entire region by a plasmid bearing the relevant fragment (see text). Only those fragments discussed in the text or used for sequencing are shown here. Boxes below the restriction map show the locations of sequenced genes (20; this report); dashed arrows show the direction of translation.

MATERIALS AND METHODS

Starting plasmid and phenotype tests used. In the work presented here, we examined the properties of our clone, pER105 (48), in isogenic strains related to the parental strain. This clone was isolated by using specific selection for ability to restrict T4 with ^{hm5}C in its DNA (T4α*gt57*β*gtl4*). It is pBR322 carrying an 8-kb chromosomal *Bam*HI fragment (Fig. 1) from our wild-type strain, ER1370, and was shown to complement seven *mcrB* mutations by several different tests for restriction activity (48). These included restriction of phage λ that had been modified by M · *Msp*I (methylated sequence ^{me}CCGG [59]; hereafter called λ · *Msp*I), M · *Hae*II (recognition site RGCGCY; position of methylation unknown; hereafter called λ · *Hae*II), or M · *Dde*I (^{me}C TNAG [24]; hereafter called λ · *Dde*I) in addition to T4*gt*. McrB$^+$C$^+$ restriction in a wild-type strain reduces the titer of the first three test phage by factors of 5 to 100 relative to the titer on an McrB$^-$ control strain, while it reduces the titer of T4*gt* by a factor of 10^6 to 10^7.

Strains, phages, and microbiological procedures. All but one strain used in this work were derivatives of ER1370, which is *trp-31 his-1 argG6 rpsL104 fhuA2* Δ(*lacZ*)*r1 supE44 xyl-7 mtl-2 metB1 serB28*. Derivatives used, with additional alleles carried, were as follows: ER1564, *mcrA1272*::Tn*10 hsdR2* Ser$^+$; ER1565, *mcrA1272*::Tn*10 mcrB1 hsdR2* Ser$^+$; ER1648, Δ(*mcrBC-hsdRMS-mrr*)*2*::Tn*10 mcrA1272*::Tn*10* Arg$^+$ Ser$^+$; ER1727, which is ER1648 with F′ *lac proAB lacI*^{q} Δ(*lacZ*)*M15*; and ER1729, *mcrA1272*::Tn*10 mcrB1 hsdR2*? Arg$^+$ Ser$^+$[F′ *lac proAB lacI*^{q} Δ(*lacZ*)*M15*]. In some preliminary experiments, ER1451 (=JM107 *mcrB1*) was also used. Construction of these strains was reported previously (48) or was similar to constructions reported there. The F′ donor was 71-18 (35).

Phages used were (i) λ *vir* grown on strains with or without various cloned methylase genes and (ii) T4D and its nonglucosylated mutant derivative T4α*gt57*β*gt14* (19; hereafter called T4*gt*). In all cases, λ phage were K modified. Phage R408 (54) was used as helper for growth of transducing lysates of plasmids carrying the packaging origin of phage f1.

Media, phage growth, and phage plating experiments were as described previously (48) except that plaque counts of spot titers were routinely used for restriction tests both with modified λ and with T4*gt*. Full-plate titers were used to verify low levels of restriction (two- to fivefold). The degree of restriction observed was more reproducible when bacteria were grown with vigorous aeration to mid-exponential phase (~1 × 10^8 to 2 × 10^8 CFU/ml), as was originally observed for *mcrA* (*rglA*) (33). Kanamycin, chloramphenicol, and ampicillin were used at 20, 15, and 100 μg/ml, respectively.

Restriction of plasmid transduction. In some experiments, restriction was tested by introducing a plasmid bearing the *mcrB* gene or one bearing both *mcrB* and *mcrC* [these plasmids are referred to together as *mcrB*(*C*) below] into cells expressing a resident methylase. In these experiments,

the *mcrB(C)* plasmid carried an f1 packaging origin and could be packaged into infectious particles by R408, a helper phage partially defective in self-packaging (54). This resulted in a phage lysate capable of transducing the *mcrB(C)* plasmid into F-containing cells. After heat treatment to kill residual donor bacteria (62), such transducing lysates were cross-streaked against suspensions of actively growing cells containing various methylase plasmids on plates containing both chloramphenicol [to select for the *mcrB(C)* plasmid] and ampicillin (to select for the resident methylase plasmid). Inviable combinations (Mcr^+ with a methylase capable of creating a target site) failed to form transductant colonies; viable combinations (Mcr^- subclone or vector alone as infecting plasmid, or a resident methylase not capable of creating a target site) formed confluent stripes of transductant bacteria.

Restriction enzymes and cloning procedures. Calf intestinal alkaline phosphatase was from Pharmacia. *Mae*II was from Boehringer Mannheim Biochemicals and was a generous gift of Rich Roberts. All other restriction enzymes, Klenow fragment, T4 DNA ligase, and T4 polynucleotide kinase were from New England Biolabs and were used as recommended. Deoxynucleoside-α-thiotriphosphates were from Promega Biotec. [γ^{32}P]ATP was from Dupont, NEN Research Products. Agarose was Seakem LE or Seaplaque low-melting-point agarose (FMC Bioproducts).

Procedures for digestion, ligation, dephosphorylation, and phosphorylation of DNA and for agarose gel electrophoresis were substantially as described by Maniatis et al. (34). Plasmids were prepared by a CsCl-ethidium bromide equilibrium ultracentrifugation procedure (34) or a modified mini-cleared lysate procedure (14; treatment with diethylpyrocarbonate was omitted, and a phenol extraction was added after the heat precipitation step). For analytical purposes, restriction reactions were 2- to 10-fold overdigested in a 20-μl reaction volume; agarose gels were 0.7 or 1% and were run in Tris-acetate buffer. For fragment isolation, a pilot experiment determined more accurately (i) the appropriate amount of enzyme to use to ensure complete digestion and (ii) the DNA concentration. Then, 10 μg of miniprep or CsCl-purified DNA was digested at a concentration of 50 μg/ml. After electrophoresis (0.7% agarose run in Tris-acetate buffer), an appropriate gel slice was excised from a portion of the gel not exposed to UV light. The DNA was eluted by a freeze-thaw method (J. Benner, personal communication) or from low-melting-point agarose as instructed by the manufacturer; 30 to 70% recovery was obtained on different occasions. For construction of ΔRI* (see below), 100 ng of pER105 DNA was digested with 10 U of *Eco*RI for 4 h and recircularized. Single-stranded DNA was made according to Russel et al. (54).

Plasmid constructs. Plasmids carrying modification methylase genes were p*Msp*I 1-30 (41); pER82 (48), which carries the gene for M · *Hae*II; p*Alu*IM12.0 (T. Jaeger and G. Wilson, unpublished data); p*Hpa*IIM (C. Card and G. Wilson, unpublished data); p*Pvu*IIM1.9 (7); p*Ban*IIH3-19 (G. Wilson, unpublished data); and p*Dde*IM1.6 (24). All of these constructs are based on pBR322 and confer ampicillin resistance. Plasmid vectors for other constructs were pBR322 (8), pACYC184 (11), pEMBL19m(+) (15), and pDD34. pDD34 is pACYC184 with the intergenic region (the replication and packaging origin) from single-stranded phage f1 cloned in the *Cla*I site. It was made by in vitro deletion of the *Hin*dIII fragment carrying the *Eco*RI methylase gene from pJH16 (23), followed by recircularization of the resultant plasmid. pER105, the starting plasmid for this analysis, is an 8-kb *Bam*HI fragment (Fig. 1) in the *Bam*HI site of pBR322, with the *tet* promoter reading into the fragment from the left in Fig. 1. Other plasmid constructs are listed on Fig. 1 and described below.

Subcloning of the *mcrBC* region. Subcloning was done by using a host strain with a deletion of all of the DNA present on pER105, so that all functions relevant to restriction in various tests could be found. Standard deletion analysis began with pDD6, an *Xba*I fragment of pER105 (Fig. 1) subcloned into the high-copy-number vector pEMBL19m(+), with the fragment oriented so that the vector *lac* promoter reads into the fragment from the right in Fig. 1. This plasmid was then analyzed by using restriction sites in the polylinker and in the fragment, as described in detail below, to yield most of the eponymous deletions (e.g., Δ*Stu*I) in Fig. 1. Nested deletions beginning at the *Stu*I site were then made starting with pDD6, using *Eco*RI to cleave in the polylinker (to the left in Fig. 1), *Stu*I to cleave in the fragment, and exonuclease III (ExoIII) and mung bean nuclease to generate a series of deleted fragments, yielding the series of numbered fragments Δ4 to Δ12 shown in Fig. 1. Only constructs that preserved the vector and the right end of the insert (as in Fig. 1) intact were analyzed; thus, all of them carry the same right and left vector sequence arrangements. These were used for sequence analysis. Selected constructs were recloned into a pACYC184 derivative (see Fig. 1 and below) for verification of phenotype (see Results).

Details of the constructions were as follows. For pDD6 and -7, the *Xba*I fragment of pER105 was gel purified and ligated into the *Xba*I site of pEMB19m(+) and pACYC184, respectively, and colonies were screened for McrBC activity (restriction of T4*gt* by cross-streak). The fragment is oriented so that *mcrBC* is read with the *lac* promoter in pDD6 and against the *tet* promoter in pDD7. In the pDD6 construction, this was the only orientation obtained (21 of 21 $McrB^+C^+$ plasmids examined), whereas in the pDD7 construction, three-fourths of the $McrB^+C^+$ products were in the orientation described. pDD9 was made by ligating the *Xba*I *mcrBC* fragment of pDD6 into the *Xba*I site of pACYC184, and the fragment is in the orientation opposite that in pDD7. pDD50 was made starting with pER105 (the *Bam*HI fragment in pBR322) in three steps. First, the *Apa*I-*Stu*I fragment was deleted and replaced with an *Apa*I 8-mer linker. Second, an *Eco*RI fragment carrying the kanamycin resistance cassette from pSKS114 (58) was filled in, ligated to *Apa*I linkers, digested with *Apa*I, and ligated into the *Apa*I site created in the first step. Third, the *Bam*HI fragment from this intermediate construct was purified and ligated into the *Bam*HI site of pDD34. Two series of deletion plasmids were made from pDD6. One series was made by digesting with *Sma*I, which cuts in the polylinker (to the left above), and with one of several enzymes that cut uniquely in the *mcrBC* fragment, filling in with Klenow fragment if necessary, and recircularizing with T4 DNA ligase. This yielded several $McrB^+C^+$ plasmids of different sizes, of which Δ*Stu*I was the smallest. Digestion of pDD6 with *Pst*I (which cuts in the polylinker to the right above) and *Nsi*I, followed by digestion with T4 DNA polymerase to yield flush ends, followed by recircularization, yielded an $McrB^-C^-$ plasmid, as did digestion with *Mlu*I followed by recircularization (not shown). The second deletion series was made with ExoIII and mung bean nuclease, using the ExoIII-Mung DNA Sequencing System kit from Stratagene as directed by the manufacturer. pDD6 was digested with *Eco*RI, filled in with deoxy-α-thiotriphosphate nucleotides, and then digested with *Stu*I. The resulting fragment was gel

purified and then digested with ExoIII for various times, and samples were treated with mung bean nuclease. After phenol extraction and isopropanol precipitation, these fragments were filled in with Klenow fragment, ligated to *Not*I linkers, gel purified, digested with *Not*I, and recircularized with T4 DNA ligase. Δ12 has the same sequence as does Δ*Stu*I except for polylinker sequences between the *Eco*RI site and the *Sma*I site, which are present in Δ*Stu*I but not Δ12. Only Δ9 and Δ10 actually received a *Not*I linker. To transfer these constructs into a low-copy-number environment, the entire *mcrBC* fragments from Δ8, Δ9, Δ10, and Δ12 were isolated by digestion with *Pvu*II, which cleaves in flanking vector sequences (within the *lac* α-fragment gene and just upstream of the *lac* promoter) and not within *mcrBC*. The fragment thus released was gel purified, ligated to *Bam*HI linkers, digested with *Bam*HI, gel purified a second time, and finally ligated into the *Bam*HI site of pDD34. Plasmids in which the *mcrBC* region was in the same orientation with respect to the *tet* gene as in pER105 were chosen for analysis and are named in the pACYC column in Fig. 1. An additional deletion, ΔRI*, was made from pER105 by overdigestion with *Eco*RI (to cleave at *Eco*RI* sites), followed by recircularization. Two independent deletions were isolated in this way. Both deletions began at the vector *Eco*RI site and ended at the same *Eco*RI* site in the insert. These two constructs and Δ10 exhibited the same distinctive phenotype, designated $McrB^+C^-$ in Fig. 1 (see Results).

In vitro transcription-translation reactions. Reactions were done according to the instructions in the Prokaryotic DNA Expression System kit from Dupont, NEN with the following modifications: (i) all reaction mixtures were 10μl; (ii) reactions were incubated for 55 min at 37°C; and (iii) after 45 min, 1 μl of a 0.5-mg/ml concentration of RNase A was added to each reaction. The products of each reaction were analyzed on 10 to 20% polyacrylamide gradient gels (Integrated Separation Systems; Emprotech, Hyde Park, Mass.). Gels were run, fixed, dried, and autoradiographed according to Ausubel et al. (3) (units 10.2 and 12.1), using Kodak XAR-5 X-ray film, with exposure at −70°C. Size standards were β-lactoglobulin (18.4 kilodaltons [kDa]), carbonic anhydrase (29 kDa), ovalbumin (43 kDa), bovine serum albumin (68 kDa), phosphorylase *b* (97.4 kDa), and myosin heavy chain (200 kDa), provided as a mixture prelabeled with ^{14}C (Bethesda Research Laboratories, Inc.).

DNA sequencing and data acquisition. Dideoxy DNA sequencing (56) was performed as described by Williams et al. (61) or by a modification of the collapsed-plasmid protocol (21, 21a, 56). All reactions used [α-^{35}S]dATP (600 to 1,200 Ci/mmol; Dupont, NEN) in conjunction with the DNA polymerase I (large) Klenow fragment or *Thermus aquaticus* (*Taq*) DNA polymerase and premixed [^{35}S]DNA sequencing reagents (New England Biolabs) and were electrophoresed on 6 to 8% standard, wedge, or buffer gradient sequencing gels (2, 42). Both strands were completely sequenced. When compressions were found, sequencing was also done at 55°C, using *Taq* DNA polymerase. Oligonucleotide primers were synthesized on a BioSearch 8600 automated synthesizer and gel or high-performance liquid chromatography purified (New England Biolabs, Organic Synthesis Division). Data were entered and analyzed by using a GrafBar Digitizer (SAC Corp.) in conjunction with computer programs from Cold Spring Harbor Laboratory (25) or the University of Wisconsin Computer Group (UWGCG; 17).

Sequence analysis was done with programs from UWGCG (17). Figure 3 was generated by using the program PUBLISH, followed by hand editing. Figure 5 was generated by using the program WINDOW, with a window of 200, a shift of 10, and a sequence motif of s (=G+C), and displayed with STATPLOT. Net charge and molecular weights were derived by the program PEPTIDESORT. Data base searching for specific motifs was done with FIND, and searching for overall similarity was done with WORDSEARCH and SEGMENTS. The latter yielded no proteins more similar to McrB or -C than expected at random, when evaluated according to the procedure recommended by the UWGCG package. Particular protein sequences that might be related to McrB or McrC were selected for functional reasons, such as possession of nuclease activity or for possession of particular structural motifs, such as a helix-turn-helix motif. These were then investigated in more detail by using GAP, BESTFIT, and COMPARE (displayed with DOTPLOT). No similarities of interest were found. Dyad symmetries were detected with STEMLOOP.

Dyads in Fig. 6 were aligned by hand at the centers of symmetry. To evaluate the significance of the similarity observed, similar pairwise alignments were made of dyad E with seven factor-independent terminators obtained from reference 46, chosen for similarity of base composition, and identical bases were scored as shown in Fig. 6. The percentage identity was as expected from the base composition. This value was calculated from the expectation: the probability that the next base in both sequences is, for example, an A is the product of the probabilities that the next base in each sequence is an A. Thus, the probability of identity is the sum of the squares of the frequencies of each base. This sum is 0.253 in this sample.

Nucleotide sequence accession number. The sequence reported here has been assigned GenBank accession number M34235.

RESULTS

The McrB sequence specificity is the sum of two parts. Our original 8-kb *Bam*HI clone, pER105, was subcloned (Fig. 1) in three different plasmid vectors as described in Materials and Methods. These subclones were analyzed for ability to complement a large deletion of the chromosomal restriction region, Δ(*mcrBC-hsdRMS-mrr*)*2*, in tests of restriction of phage with modification patterns foreign to *E. coli*. The deletion strain used was ER1727. In this situation, complementing plasmids clearly fell into two classes (Table 1). The larger insert fragments, Δ12 and *Bam*HI (Fig. 1), conferred restriction competence in all three restriction tests used (lines 3 and 4), but the smaller fragments, Δ10 and ΔRI*, conferred restriction competence only when λ · *Msp*I was the test phage (lines 1 and 2). The smaller plasmids conferred no restriction competence whatever when the test phage was λ · *Hae*II or T4*gt*. We infer that these smaller fragments encode a function (McrB) able to restrict λ · *Msp*I by itself, whereas the larger fragments encode as well a function (McrC) that confers the ability to restrict λ · *Hae*II and T4*gt* when combined with McrB. The requirement for McrC was absolute; if McrB alone restricted T4*gt* with reduced efficiency, we should have been able to distinguish this from no effect at all, since the maximum effect was very large: less than one phage in 10^6 formed plaques when the cell contained the Δ12 or *Bam*HI fragments (lines 3 and 4) or the chromosomal wild-type allele (line 13). This was true regardless of whether the vector was pBR322 (lines 2 and 4) or pACYC184 (lines 1 and 3). The different vectors were used to protect against vector-specific or construct-specific artifacts (discussed below).

The *mcrB1* allele inactivates McrB completely, but McrC

TABLE 1.

Plasmid	Fragment	Plasmid origin	Host strain	Plating efficiency[a]			Inferred genotype (plasmid/chromosome)
				λ · *Msp*I	T4*gt*	λ · *Hae*II	
1. pDD39	Δ10	pACYC	ER1727	**0.13**	1.2	0.7	$B^+C^-/\Delta BC$
2. pDD11	ΔRI*	pBR322	ER1727	**0.1**	0.8	0.7	$B^+C^-/\Delta BC$
3. pDD41	Δ12	pACYC	ER1727	**0.13**	**$<10^{-6}$**	**0.1**	$B^+C^+/\Delta BC$
4. pER105[b]	*Bam*HI	pBR322	ER1727	**0.05**	**$<10^{-6}$**	**0.05**	$B^+C^+/\Delta BC$
5. pDD34	None	pACYC	ER1727	0.8	0.8	0.7	$B^-C^-/\Delta BC$
6. pBR322	None	pBR322	ER1727	1.1	1.3	1.0	$B^-C^-/\Delta BC$
7. pDD39	Δ10	pACYC	ER1729	**0.09**	**0.09**	**0.5**	$B^+C^-/B^-C^{limiting}$
8. pDD11	ΔRI*	pBR322	ER1729	**0.02**	**0.001**	**0.5**	$B^+C^-/B^-C^{limiting}$
9. pDD41	Δ12	pACYC	ER1729	**0.05**	**$<10^{-6}$**	**0.04**	$B^+C^+/B^-C^{limiting}$
10. pER105	*Bam*HI	pBR322	ER1729	**0.03**	**$<10^{-6}$**	**0.02**	$B^+C^+/B^-C^{limiting}$
11. pDD34	None	pACYC	ER1729	0.8	1.2	0.8	$B^-C^-/B^-C^{limiting}$
12. pBR322	None	pBR322	ER1729	1.0	1.5	1.1	$B^-C^-/B^-C^{limiting}$
13. None	None	None	ER1564	**0.03**	**$<10^{-6}$**	**0.04**	B^-C^-/B^+C^+

[a] Ratio (titer of the phage on strain X)/(titer of the phage on permissive host ER1565). The value for control phage (λ *vir* · K or T4D) was 0.8 to 1.2 in all cases. Numbers in bold represent restriction-positive combinations of plasmid, host strain, and test phage.
[b] Done in a separate experiment.

function remains partially active. Also in Table 1 we characterize the *mcrB1* allele, the one found in such common laboratory strains as WA802 (K802), MC1061, and ED8767 (47, 48). This mutation is known to be either a point mutation or a very small multibase alteration (48). The strain used here, ER1729, is otherwise isogenic with the deletion strain.

The McrB function was inactivated by *mcrB1*, since no restriction was seen in any test, even λ · *Msp*I, when the vector alone was present (Table 1, lines 11 and 12). Restriction of McrB-specific λ · *Msp*I was completely restored by all four fragments tested (lines 7 to 10).

The chromosomally encoded McrC function was not inactivated by the *mcrB1* mutation. Restriction of λ · *Hae*II and T4*gt*, which requires McrC (see above), was restored by the Δ10 and ΔRI* fragments (Table 1, lines 7 and 8), although restoration was only partial. For T4*gt* the reduced level of restriction is still substantial, but since the original restriction of λ · *Hae*II was only 10- to 50-fold (lines 9, 10, and 13), the residual 2-fold restriction was unimpressive (lines 7 and 8) but reproducible (data not shown). On the other hand, the Δ12 and *Bam*HI fragments restored restriction to wild-type levels (compare lines 9 and 10 with line 13).

The results also demonstrate that McrB is still required for restriction T4*gt*, although it is not sufficient alone. If McrC alone were sufficient for this restriction, then restriction by an *mcrB1* strain should have been the same whether the plasmid carried no fragment (lines 11 and 12) or an *mcrB*-only fragment (lines 7 and 8). For T4*gt*, at least, these values clearly differ, so McrB is required.

One reasonable interpretation of these results is that the *mcrB1* mutation reduces the level of McrC produced, for example via a polar effect, but does not eliminate it entirely. McrC is thus limiting for restriction activity in this situation. The fragments that carried *mcrB* alone were unable to restore restriction to wild-type levels because of this limitation, but some restriction was observed, implying that some level of each wild-type protein was synthesized.

Sequences sensitive to restriction by McrB or McrBC. We characterized further the sequence specificity of the two restriction functions by introducing the $McrB^+C^-$ or $McrB^+C^+$ plasmids into cells already carrying a cloned methylase gene, which is a lethal situation if restriction occurs (see Materials and Methods for details of the test). The methylases in question modify eight different sequences; the sequences and positions of modification (Table 2) are from the indicated references: *Msp*I, *Hae*III, *Alu*I, and *Hpa*II (59); *Alu*I and *Pvu*II (10); and *Dde*I (24). The results suggest that whereas $McrB^+C^+$ restriction acts on many methylated sites, McrB restriction is limited to a small number of sites. The plasmid bearing the gene for M · *Msp*I is the only one of eight tested that caused sensitivity to McrB independently of McrC. McrBC, on the other hand, recognized sites created by seven of the eight methylases tested, and this restriction specificity was thus quite relaxed. Three different cytosine modifications can confer sensitivity to McrBC: methylation at the C5 position (e.g., M · *Alu*I), methylation at the N4 position (M · *Pvu*II), and hydroxymethylation at the C5 position (T4*gt*). Since McrB alone apparently does not recognize ^{hm}C (Table 1), it may be that McrC confers the ability to recognize additional modifications as well as additional sequence contexts.

The magnitude of restriction depends on the vector used. During manipulations designed to facilitate genetic and physical analysis of the *mcrBC* region, we found that use of pEMBL19 (a pUC derivative) as vector severely reduced the magnitude of restriction. Table 3 gives representative results for restriction of three different methylated targets. As

TABLE 2. Sequences sensitive to McrB and McrBC restriction

Methylase	Modified sequence[a]	Restricted[b] by Δ(*mcrBC*)2 + plasmid carrying:	
		mcrB	*mcrBC*
*Msp*I	^{m5}C CGG	+	+
*Hae*III	GG ^{m5}C C	−	+
*Hae*II	RG ^{m}C GCY or		
	RGCG ^{m}C Y	−	+
*Dde*I	^{m5}C TNAG	−	+
*Pvu*II	CAG ^{m4}C TG	−	+
*Ban*II	GRG ^{m}C YC or		
	(GRGCY ^{m}C)	−	+
*Alu*I	AG ^{m5}C T	−	+
T4^{hm}C	N ^{hm5}C N	−	+
*Hpa*II	C ^{m5}C GG	−	−

[a] R, Purine; Y, pyrimidine; ^{m4}C, N^4-methylcytosine; ^{m5}C, 5-methylcytosine; ^{hm5}C, 5-hydroxymethylcytosine; ^{m}C, methylated or possibly methylated but the position of the methyl group on the base is not known.
[b] +, Restriction is observed; −, restriction is not observed. *Dde*I-modified λ and T4*gt* were tested for phage plating efficiency as in Table 1. All other methylases were tested by plasmid transduction as described in Materials and Methods.

TABLE 3. Vector dependence of magnitude of restriction

Chromosome allele[a]	Plasmid[b]		Inferred *mcrBC* copy no.[c]	Plating efficiency[d] of:		
	Vector	*mcrBC* fragment		λ · *Msp*	T4*gt*[e]	λ · *Hae*II
1. Δ(*mcrBC*)2	pEMBL19	*Xba*I (6 kb)	~200–300	**0.31**	1	0.8
2. Δ(*mcrBC*)2	pBR322	*Bam*HI (8 kb)	~30–60	**0.06**	**$<10^{-6}$**	**0.11**
3. Δ(*mcrBC*)2	pACYC184	*Xba*I (6 kb)	~15	**0.08**	**$<10^{-6}$**	**0.11**
4. Wild type	None	None	(1)	**0.008**	**$<10^{-6}$**	**0.17**
5. Δ(*mcrBC*)2	pACYC184	None	0	1.1	1	0.8
6. Δ(*mcrBC*)2	None	None	0	0.8	1	0.6
7. *mcrB1*	None	None	0	(1)	(1)	(1)

[a] Strains used were: ER1648 (lines 1 to 3, 5, and 6), ER1564 (line 4), and ER1565 (line 7). These are isogenic strains differing only in the *mcrB* allele carried. Similar results were also obtained in ER1727 and in the *mcrB1* strain ER1451.

[b] Plasmids used were pDD6 (line 1), pER105 (line 2), pDD7 (line 3), and pACYC184 (line 5).

[c] Copy number per chromosome as cited in the text, with copy number of the chromosome 1 by definition. We did not rigorously measure copy number, but the DNA yield in plasmid preparations was consistent with the relative abundance.

[d] Ratio of (titer of the phage on strain X)/(titer of the phage on ER1565). Plating efficiency on ER1565 (line 7) is 1 by definition (indicated by parentheses). The value for unmodified λ was 0.8 to 1.1 in all cases. Numbers in bold represent restriction-positive combinations of plasmid, host strain, and phage.

[e] Done in a separate experiment.

before, the host strain carried a deletion covering all of the cloned DNA. The copy numbers per chromosome of the different plasmids (and therefore of the *mcrBC* region) was taken from the indicated references: pEMBL19 (36); pBR322 (32, 36); and pACYC184 (11). When the cloned DNA was carried on low-copy-number pACYC184 or pBR322, restriction was comparable to restriction by a chromosomal copy of the wild-type genes (Table 3; compare lines 2 and 3 with line 4). On the other hand, restriction was almost undetectable in these tests when the same *Xba*I fragment used in the pACYC184 construct was carried by very high copy number pEMBL19m (line 1). Essentially identical results were obtained with a strain carrying the *mcrB1* mutation (not shown).

We have eliminated several possible explanations for these results. First, plasmid loss does not account for the reduction in restriction, since for all plasmids, colonies formed by plasmid-containing cultures on drug-free plates were drug resistant when replica plated to drug plates. The failure of the pEMBL construct to mediate restriction is not dominant, since wild-type ER1564 restricted normally whether carrying the pEMBL vector alone, an $McrB^+C^+$ subclone (pEMBLΔ*Stu*I; pDD17), or an $McrB^-C^-$ subclone (pEMBLΔ5 [pDD24] or pEMBLΔ8 [pDD27]) (data not shown). This observation rules out explanations that depend on copy number per se. Although the effect is *cis* specific, it is not a trivial result of local sequence arrangements at the vector-insert junction. Since recloning a fragment carrying about 100 base pairs (bp) of pEMBL19 sequence on either side of the Δ12 ($McrB^+C^+$) insert from nonrestricting pDD31 (data not shown) into pACYC184 to yield pDD41 restored restriction (Table 1). The effect is not the result of a requirement for transcription originating at a plasmid promoter, since in the active constructs the open reading frames (see below) can be in either orientation relative to plasmid promoters (not shown), whereas in the inactive pEMBL constructs they are in the same orientation as the strong *lac* promoter.

In any event, the results shown strongly suggest that a negative result with one vector plasmid should be interpreted with caution, since a different result might be obtained in a different vector.

In vitro transcription-translation of the *mcrBC* region. The genetic evidence presented above suggests a minimum of two components of restriction. As described in detail below, in vitro transcription-translation experiments (Fig. 2) showed that four polypeptides were synthesized under the direction of the *mcrBC* fragment, of sizes roughly 49, 40, 38, and 29 kDa. The 29-kDa and the 49-kDa polypeptides appeared to be produced from the same reading frame, designated the *mcrB* frame. The 40-kDa and the 38-kDa polypeptides appeared together when the plasmid carried *mcrC*, with the 40-kDa polypeptide expressed at much reduced efficiency compared with the 38-kDa polypeptide. There is insufficient coding capacity to account for all four of these proteins, necessitating an explanation invoking internal translational initiation or product processing.

Two *mcrB*-specific proteins, about 49 and 29 kDa, were visualized when synthesis was programmed with $McrB^+C^+$ or $McrB^+C^-$ plasmids. Δ12 ($McrB^+C^+$) showed proteins of 29, 38, 40, and 49 kDa (Fig. 2B, lane 13, arrows a to d respectively). The 29-kDa band was not β-lactamase, since the *bla* gene had been interrupted by digestion with *Fsp*I, and β-lactamase was missing (Δ4, lane 10; Δ9, lane 11) or greatly reduced (lane 9, pEMBL) in the controls. Synthesis of the 49- and 29-kDa but not the 38- and 40-kDa proteins was also programmed by *Fsp*I-cut Δ10 ($McrB^+C^-$; lane 12, arrows a and d). Both of these proteins are therefore encoded by the *mcrB* gene, as defined in the foregoing genetic analysis.

The 38- and 40-kDa proteins seen in Δ12 (best visualized in Fig. 2A, lane 6, arrows d and e) were the products of the *mcrC* gene. Neither of the $McrB^+C^-$ plasmids synthesized either intact protein: ΔRI* did not synthesize such a protein at all (lane 5), and although Δ10 did synthesize a protein of 38 kDa in the absence of *Fsp*I digestion (lane 4), this protein disappeared with digestion (Fig. 2B, lane 12), presumably because the *Fsp*I site in the *lacZ*α gene was interrupted by digestion. This result is consistent with sequence data (below and not shown), which show an in-frame fusion of *mcrC* to the *lacZ*α gene in this construct.

The suggestion that two products, of 29 and 49 kDa, are synthesized from the *mcrB* gene is supported by the fact that Δ9 appeared to program synthesis of two proteins sensitive to digestion of the *lacZ*α gene. Without digestion, three proteins, of 29 (β-lactamase), ~55, and ~37 kDa were synthesized (Fig. 2A, lane 3); all of these were abolished by *Fsp*I digestion (Fig. 2B, lane 11). There are only two *Fsp*I sites in the plasmid, one in the *bla* gene and one in the *lacZ*α gene; digestion of one of these interrupts *bla*, so the other must be interrupting the coding sequence for two protein products. Sequence data (see below) show an in-frame fusion of *mcrB* to the *lacZ*α gene, yielding a frame sufficient to code for a protein of 55 kDa, like the one observed. The

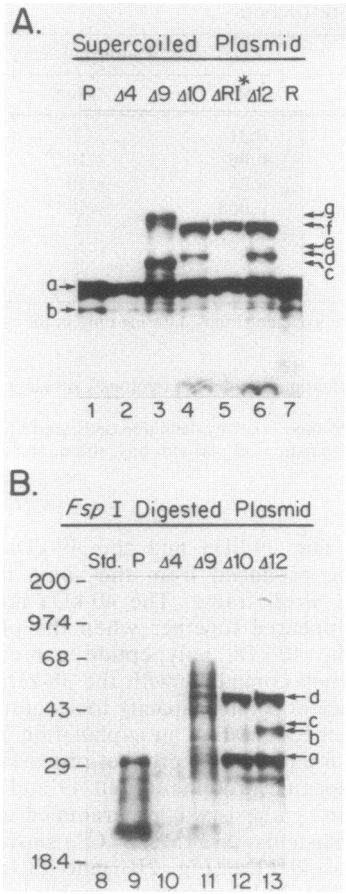

FIG. 2. Identification of the McrB and McrC polypeptides. In vitro transcription-translation reaction products were analyzed on sodium dodecyl sulfate-polyacrylamide gels. Plasmids examined carried the fragments Δ12 (pDD31; McrB$^+$C$^+$), Δ10 and ΔRI* (pDD29 and pDD11; McrB$^+$C$^-$), and Δ4 and Δ9 (pDD23 and pDD28; McrB$^-$C$^-$). All except ΔRI* are carried in pEMBL19m. pEMBL19m (P) and pBR322 (R) were included as controls. Reactions were programmed with uncut supercoiled plasmid DNA (A) and *Fsp*I-digested plasmid DNA (B). Digestion of the DNA reduced (lane 9) or eliminated (lanes 10 and 11) synthesis of β-lactamase and of fusion proteins joining the *lac*α fragment in frame to McrB (Δ9; lane 11) or McrC (Δ10; lane 12). Arrows in panel A: a, pre-β-lactamase; b, β-lactamase; c, product of fusion to small McrB protein from Δ9; d, proposed second initiation product of *mcrC*; e, McrC and product of fusion to McrC from Δ10; f, McrB; g, product of fusion to large McrB protein from Δ9. Arrows in panel B: a, small McrB protein; b and c, small and large McrC proteins; d, large McrB protein. Molecular sizes calculated from gels like this (with conversion for about 10% fast migration of the size standards in parentheses) are as follows: large McrB protein, 49 kDa (53.2 kDa); large McrC protein, 40 kDa (43.4 kDa); small McrC protein, 37.5 kDa (40.7 kDa); small McrB protein and β-lactamase precursor, 29 kDa (31.5 kDa). β-lactamase precursor is actually 31.5 kDa, according to the sequence in the NBRF data base.

~37-kDa protein cannot result from truncation or fusion of the *mcrC* reading frame, which has been entirely deleted, so it must result from translation of the first open reading frame. Thus, two proteins are read in the same frame and are disrupted by the same restriction digest.

Our results are broadly consistent with those of Ross et al. (51, 52) except that we detect a doublet, rather than one band, in the 38- to 40-kDa region.

Sequence of the *mcrBC* region and sequence verification. To clarify the relationships among all of the polypeptide products and between these and the genetic data, we sequenced the *Hpa*I-*Stu*I region as well as the junctions between *mcrBC* sequence and the vector in the subclones used above. While this paper was in preparation, another version of the sequence appeared (53). There are substantial differences between that sequence and ours, so we present it in full in Fig. 3. Our sequence differs at 15 positions: 14 nucleotides are inserted and 1 is deleted relative to the published sequence (53), resulting in a net gain of 13 nucleotides, all but 1 within a region of about 250 bp (boxed with dashed lines in Fig. 3).

We have sequenced both strands, and at each position of disagreement with the sequence of Ross et al. (53) the sequence was determined from three different priming sites, as was 80% of the entire sequence. To determine whether natural sequence variation might exist in this region, we determined the sequence of this region on both strands from another clone, derived from CR63 (pBg6 [55]). The sequence obtained was identical to that shown in Fig. 3, which originated in ER1370. The clone sequenced by Ross et al. (53) also derived from CR63 (55).

Of the 14 nucleotide differences found within the dashed box in Fig. 3, 11 result in differing restriction site positions for five restriction enzymes. To verify our sequence, we performed the digests shown schematically in Fig. 4B, with results in Fig. 4A, on an 857-bp *Hin*cII-*Hin*dIII fragment just downstream of *hsdS*. For each digest, line a in Fig. 4B represents the restriction map predicted by our sequence, and line b the restriction map predicted by Ross et al. (53).

Results of the digests (Fig. 4A) confirmed our sequence. Those simplest to interpret are the *Taq*I digest (lane 1) and the *Mae*II digest (lane 3), in which the prediction is either no site or one site approximately in the middle of the fragment. As predicted by our sequence, *Taq*I does not cut the fragment, and *Mae*II cuts it into two fragments of predicted sizes. The absence of the *Taq*I site is due to the insertion of two nucleotides (Cs at nucleotides 688 and 690) in our sequence that destroys the site present in the sequence of Ross et al. (53); the presence of the *Mae*II site is due to the insertion of three nucleotides (ACG at 776 to 778) in our sequence that creates the site not present in the sequence of Ross et al. One nucleotide deletion in our sequence (a G between our nucleotides 572 and 573) creates a second *Bsm*AI site; the diagnostic fragments (482 and 160 bp) can be seen in lane 9 of Fig. 4A. The most complicated case is the *Hga*I digest. Ross et al. have two sites not present in our sequence, and our sequence has one site not present in theirs. The difference results from five nucleotide insertions in our sequence relative to theirs (C at 654 abolishes their site; G at 787 abolishes another; CGT at 695 to 697 creates ours). Lane 6 of Fig. 4A shows two of the three fragments predicted by our sequence; the second-largest band is clearly between 270 and 310 bp, not 252. The 69-bp fragment was not resolved here, but in other experiments (not shown) this fragment was detected. One *Hga*I site in each sequence is an *Mlu*I site as well, but since the *Hga*I sites are in different positions, so are the *Mlu*I sites; we verified that the *Mlu*I site is located in the position predicted by our sequence rather than that predicted by Ross et al. with an *Mlu*I-*Hpa*II double digest (Fig. 4A, lane 12); the predicted 292-bp fragment, rather than a 321-bp fragment, is obtained. This result reconfirms four of the five sequence differences tested by the *Hga*I digest. Three nucleotide insertions in this region of our sequence (Cs at 714 and 749 and a T at 792), as well as one downstream of *mcrC* (a C at 2467) remain untested.

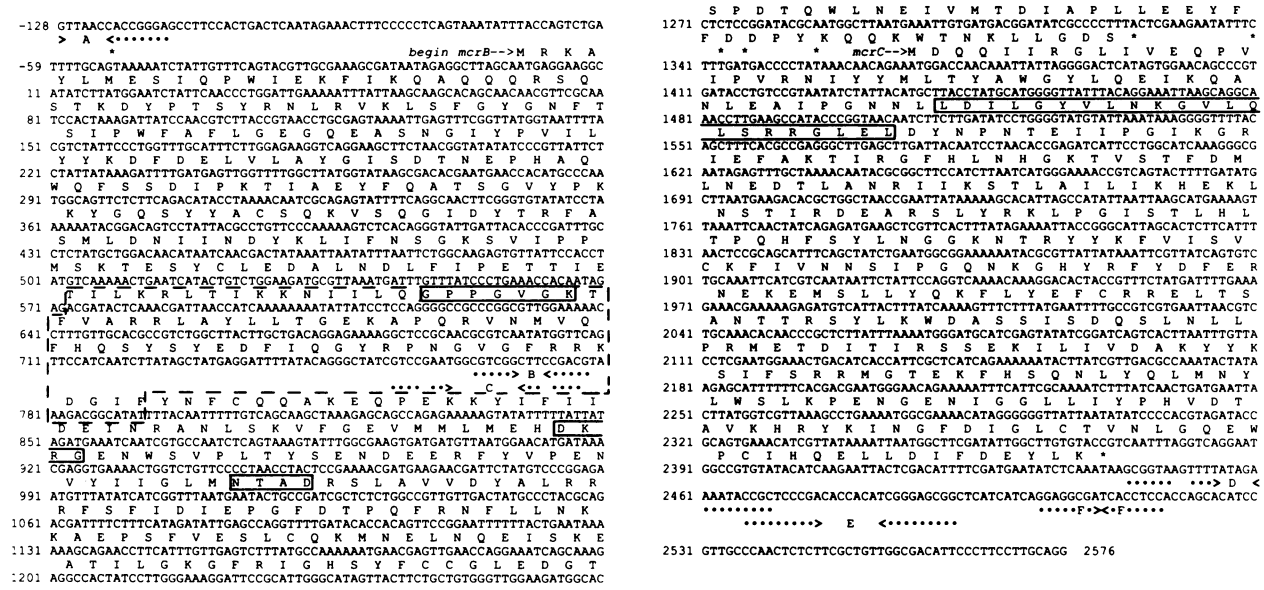

FIG. 3. DNA sequence of the *Hpa*I-*Stu*I fragment carrying *mcrBC* and protein sequences encoded by it. Base pair 0 is the A of the first ATG of the mcrB reading frame; bp -128 is the first base pair of the *Hpa*I site, and bp 2576 is the third base pair of the *Stu*I site. The orientation of the sequence is opposite that shown in Fig. 1 so that protein sequence may be read left to right. Single-letter abbreviations for amino acids are placed above the first base of the corresponding codon, and translation begins with the first ATG in each frame; we do not have evidence to rule out other starts. Notations: *, Stop codons surrounding the long open reading frames; ···>, <···, inverted repeat sequence, with a dot under each base that can pair in a hairpin. A dashed box surrounds the region (bp 574 to 795) of the sequence that disagrees with the sequence of Ross et al. (53) at 14 positions. Boxes surrounding protein sequence alone indicate protein structure motifs discussed in the text.

Inferred protein sequences. Translation of the two open reading frames observed in the sequence in Fig. 3, beginning with the first methionine of each frame, yields products of 53.9 (*mcrB*) and 41.7 (*mcrC*) kDa. Note that the orientation of the region is inverted relative to that in Fig. 1 so that the polypeptide sequence can be read forward. These sizes agree reasonably well with the sizes of the two largest proteins observed in in vitro transcription-translation experiments (49 and 40 kDa), but potential starts at nucleotides 18 (*mcrB*) and 1397 (*mcrC*) are also consistent with the observed sizes. It is possible that both potential *mcrC* starts are used, since we observe a doublet at 40 and 38 kDa (see above). The potential start at 1397 overlaps by one base the stop codon of the *mcrB* frame (UAGTG), a situation that can lead to increased efficiency of translation initiation (57).

The positions of deletion endpoints in the plasmids used to characterize the proteins are consistent with these assignments. The deletions conferring an McrB$^+$C$^-$ phenotype end within the *mcrC* frame (nucleotides 1909 for ΔRI* and 2249 for Δ10), and the one conferring an McrB$^-$C$^-$ phenotype ends within the *mcrB* frame (nucleotide 1355 for Δ9). Sequencing of the deletion junctions in Δ9 and Δ10 confirmed that the *mcrB* and *mcrC* frames, respectively, are joined in frame to the *lacZ*α gene.

A potential GTP-binding site was found within the McrB sequence. The GTP-binding motif comprises three parts: GXXXXGK—40 to 80 amino acids—DXXG—40 to 80 amino acids—NXXD (16). These three elements are boxed in Fig. 3 (following nucleotides 571, 851, and 991). The overall length of the region containing the three elements of the motif (133 amino acids) fits well with the conserved overall length in known GTP-binding proteins (80 to 160 or ~190 to 225 amino acids [16]). Although the spacing between elements (91 amino acids between elements I and II and 29 amino acids between II and III) does not match precisely the consensus spacing (40 to 80 for each), spacing was found to vary somewhat among classes of known GTP-binding proteins (16). The nucleotide sequence changes verified above result in substitution of a 74-amino-acid block (dashed box in Fig. 3) for a different 70-amino-acid block in the sequence of the McrB polypeptide deduced by Ross et al. (53). The first element of the GTP-binding site is not present in the sequence deduced in reference 53.

The McrC protein is rather basic (net charge of +8; 15 mol% H+K+R), consistent with the idea that it binds to DNA. In addition, a potential leucine heptad repeat ("leucine zipper"; $LX_6LX_6LX_6L$) was found within the protein sequence of McrC (boxed in Fig. 3). This region is followed by a particularly cationic region (7 of 20 amino acids basic, with 1 acidic; encoded by nucleotides 1613 to 1670). The heptad repeat has been associated with dimerization or protein-protein recognition functions (29), as could be the case here.

We also looked for, but did not find, helix-turn-helix (44) and zinc finger (5, 37) motifs, both of which can play roles in DNA-protein recognition. No similarity was found to any other protein in the NBRF data base, nor was any more limited similarity observed to potentially related nucleases, GTP-binding proteins, DNA-binding proteins, or transcriptional activators, evaluated as described in Materials and Methods.

Features of the nucleotide sequence. The base composition of the 2,705 bases sequenced here is 40% G+C, unusually low for *E. coli* (50% G+C). We looked more closely at the composition of the catenated *hsdRMS-mcrBC* sequence (20, 31; see above). Figure 5 shows the G+C composition of this entire assembly as a moving average. The *hsdRM* genes reflect the G+C content of *E. coli* as a whole (50% G+C), but at the beginning of *hsdS* the G+C content drops precipitously, and the remaining genes are all substantially below the average. The sequence coding for *hsdS* has an overall G+C content of 40%; *mcrB* has 40%, and *mcrC* has 37.5%.

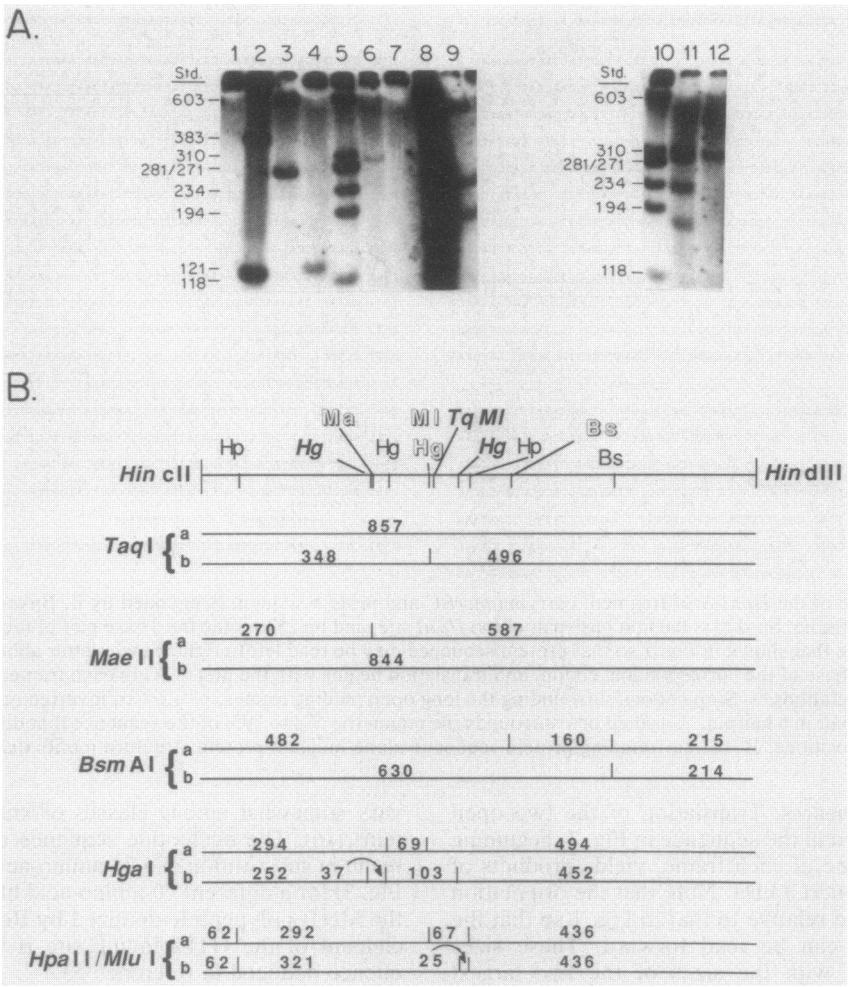

FIG. 4. Verification of DNA sequence. (A) Restriction digests used to verify sequence at disagreements. Fragments (an experimental 857-bp *Hin*cII-*Hin*dIII and a control 1,121-bp *Hin*cII fragment) eluted from agarose gels were digested with the indicated restriction enzyme, 5′ end labeled with [γ-^{32}P]ATP, and run on a 6% polyacrylamide gel. Predicted fragment sizes for the experimental lanes (1, 3, 6, 9, and 12) are shown in panel B. Size standards: lane 2, pBR322 digested with *Bst*NI (383- and 121-bp standards); lane 5, øX174 digested with *Hae*III (603-, 310-, 281-, 271-, 234-, 194-, and 118-bp standards); lane 9, pBR322 digested with *Msp*I (overexposed here). The control 1,121-bp fragment should yield subfragments of 992 and 129 bp when digested with *Mae*II (lane 4) and of 311, 294, 222, 166, 65, and 63 bp when digested with *Mlu*I and *Hpa*II (lane 11); *Hga*I should not cut it (lane 7). Fragments of less than 100 bp were not resolved in this gel. (B) Sizes predicted from sequence data for digests of the 857-bp fragment. At the top is a map of positions where sequence disagreement between Ross et al. (53) and this paper (open letters) creates a different predicted restriction enzyme cut site. Map is oriented as in Fig. 1. MI, *Mlu*I; Hp, *Hpa*II; Hg, *Hga*I; Tq, *Taq*I; Ma, *Mae*II; Bs, *Bsm*AI. Symbols in solid roman type are invariant. Below the map are shown predicted fragment sizes for specific enzymes for sequence reported here (a) and by Ross et al. (53) (b).

The G+C content rises to 47.5% in the 700-bp intergenic region between *hsdS* and *mcrB* and to 46% in the short (130-bp) region following *mcrC*. *hsdS* contains two peaks of higher (more normal) G+C content that coincide with the central and distal regions conserved at the protein sequence level in the K family of type I restriction-modification systems (20). This sequence arrangement may be a clue to the evolutionary history of the region (see Discussion).

Six elements of dyad symmetry were found strategically placed in the *mcrBC* sequence (Fig. 3; dotted arrows below the sequence). Four of the six dyads surround *mcrBC* and might represent factor-independent transcriptional terminators, which generally consist of a G+C-rich hyphenated dyad followed (usually, but not always) by a string of T residues (46). One repeat is upstream (repeat A; −134 to −113, an eight-base perfect repeat stem with six GC pairs and a five-base loop), and three are clustered immediately downstream of *mcrBC* (repeats D, E, and F; 2444 to 2469, a 10-base stem with four GC pairs, 2 unpaired bases, and a 4-base loop; 2466 to 2492, a 10-base perfect stem with eight GC pairs and a 7-base loop; and 2503 to 2518, an 8-base stem with one mismatch, five GC pairs, and no loop). The calculated free energies of these dyads are substantial: −17.8, −11.6, −26, and −10.4 kcal, respectively (1 kcal = 4.184 kJ) (60). These energies are consistent with a functional role in vivo.

Two of these dyad elements, A and E, were strikingly similar to each other and to an element located between the *hsdR* and *hsdM* genes, a region known to contain a transcription termination site(s) (31). These elements are indicated in Fig. 5 and shown in Fig. 6, aligned at their centers of symmetry. Dyad A is similar to dyad E at 15 positions; dyad A is similar to the dyad after *hsdR* at 11 positions; and dyad E is similar to the dyad after *hsdR* at 14 positions. From

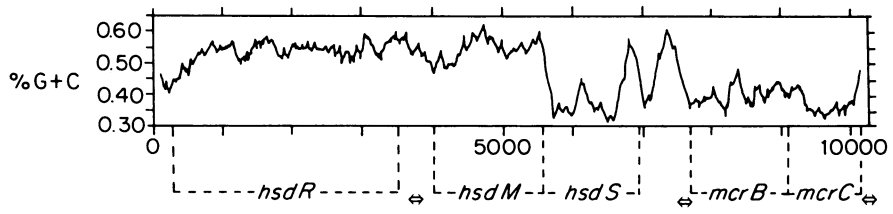

FIG. 5. Base composition of the *hsdRMS mcrBC* region. The percent G+C content of the sequence is presented as a moving average over a window of 200 bp in shifts of 10 bp at a time, generated by the program WINDOW and displayed by the program STATPLOT (UWGCG). The positions of the five coding sequences are indicated, as are the positions of three similar inverted repeats (arrows; see Fig. 6). Inverted repeats are not represented to scale.

the base composition of the set of sequences compared, we would have expected 7 to 8 matches in each comparison; indeed, alignment of dyad E with seven other dyad elements with similar base composition (see Materials and Methods) yielded 5 to 11 similar positions (average, 7.5). The similarity of sequence thus seems not to be merely a consequence of similarity of structure or base composition.

The remaining two symmetry elements overlap at one end of the region of sequence disagreement (Fig. 3, dotted arrows B and C) and might be responsible for the sequence discrepancy. Repeat B comprises a 6-base perfect repeat (five GC) with a 3-base spacer ($\Delta G = -7.8$ kcal); repeat C comprises an 8-base stem (four GC) with one mismatch and a 10-base spacer ($\Delta G = -1.0$ kcal). The two repeats share a stem, so that only one of the two structures could form at a time. We observed compressions in the region of sequence disagreement that could be accounted for by such a secondary structure.

DISCUSSION

McrBC acts like a multisubunit restriction enzyme. A consistent picture of the restriction system encoded by the *mcrBC* locus emerges when sequence evidence is added to the genetic evidence (Fig. 7). We find a two-component restriction system. One protein, McrB, is an active restriction moiety, capable of action independently of or together with the other, McrC (Table 1). McrB, which acts independently of McrC on some sites containing 5-methylcytosine (^{m5}C; Table 1), contains a protein sequence similar to known GTP-binding motifs (Fig. 3 and text), suggesting a role as a cofactor-binding subunit. The second protein, McrC, a basic protein with a potential dimerization motif (the leucine zipper; Fig. 3), interacts with McrB in some way to allow recognition not only of additional ^{m5}C-containing sites but also of additional cytosine modifications (^{hm5}C and ^{m4}C; Tables 1 and 2). However, McrC has no apparent restriction activity of its own (Table 1).

Significance of multiple protein products. Synthesis of multiple protein species from both genes (Fig. 2) in in vitro transcription-translation experiments complicates but does not contradict this picture. Synthesis of two products from *mcrB* was seen in maxicells (51–53) as well as in vitro and thus is likely to be significant. The molecular weights measured here differ slightly from those reported elsewhere (51–53), but we do not regard these as significant differences (see Materials and Methods). It was suggested that the smaller (29 kDa here) product of *mcrB* might have a regulatory role (52), but no evidence bearing on this point is available so far. In any event, there is agreement that a product of *mcrB* is required for restriction in all assays.

Synthesis of two products from *mcrC* was not reported in maxicells (52), but maxicell experiments had missed this protein altogether initially (51), so the possibility remains that

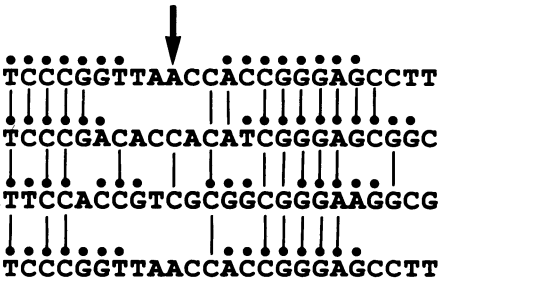

FIG. 6. Alignment of dyad symmetry elements A and E and one that lies between *hsdR* and *hsdM* (31). Arrow indicates the center of symmetry, along which the sequences are aligned. Numbering is according to the sequence presented here; that is, the *hsdRM* dyad begins 4022 bp to the left of nucleotide 0 in Fig. 3. Vertical lines connect identical nucleotides; dots over the sequence indicate positions that can pair in a hairpin. Dyad A has been represented twice so that each dyad can be aligned with both of the others.

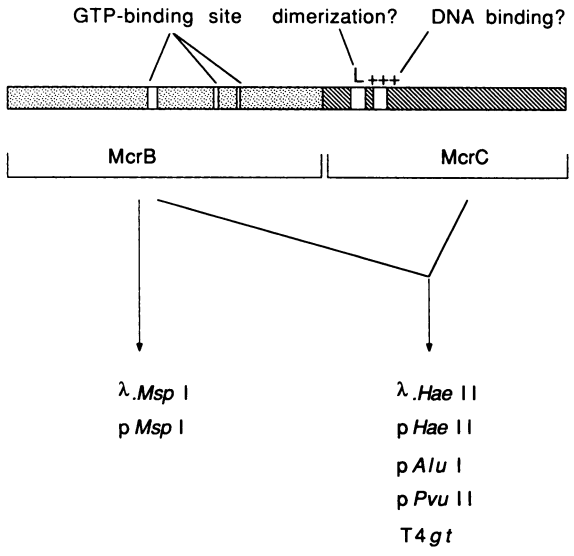

FIG. 7. Schematic picture of the McrBC restriction system. The relative sizes of the shaded boxes correspond with the sizes of the largest protein products encoded by the sequence. The positions of the sequences encoding the putative GTP-binding site, the heptad repeat (L), which could be a dimerization or other protein-protein recognition domain, and an exceptionally cationic region (+++) that could be a DNA-binding domain are indicated. Arrows indicate some of the targets restricted by McrB and McrB plus McrC.

low-level synthesis of the larger species does occur in vivo. In particular, it could account for the low level of *mcrC* complementation seen in *mcrB1* strains (Table 1), especially if translational coupling (57) accounts for predominant synthesis of the smaller species. Ross et al. have suggested both that synthesis from *mcrC* was independent of *mcrB* (52) and that it was not (53). The role of *mcrC* is in dispute (see below).

Similarity to other multisubunit restriction systems. The organization of the *mcrBC* system is reminiscent of that of type I and type III restriction systems, in which a subunit responsible for DNA binding and sequence recognition associates with one (type III) or two (type I) other subunits to constitute active enzyme (6). These systems fall into families, such that the specificity subunit of one such system can substitute for the specificity subunit of another system from the same family both in vivo and in vitro. Such reconstituted enzymes exhibit the sequence specificity of the system from which the specificity subunit is derived. Merodiploid strains carrying genes for two specificity (HsdS) subunits but only one restriction (HsdR) subunit nevertheless express both restriction specificities (9).

Such specificity subunit substitution is one way to understand the modulation of the specificity of McrB by the presence or absence of McrC (Table 2). On this model, McrC is a specificity subunit; McrB is a restriction subunit; and the two act together to restrict λ · *Hae*II, T4*gt*, and other targets. This idea is supported by the presence of a nucleotide-binding site in the McrB polypeptide sequence. The analogous subunit of *Eco*K, HsdR, similarly contains a segment proposed as nucleotide-binding site (31), and it is presumably the subunit that binds ATP. The idea is also supported by the suggestive presence of a potential protein-protein recognition motif (the leucine zipper) and associated cationic region in the McrC polypeptide sequence. This arrangement allows eucaryotic transcription factors to modulate sequence recognition by the transcription apparatus in a manner not unlike the HsdS substitution described above (1, 27, 29, 43).

Strictly speaking, the model articulated above predicts that McrB acts solely as a restriction subunit and therefore that there must exist a second specificity gene in *E. coli*, present elsewhere in the chromosome, that allows McrB to restrict λ · *Msp*I in the absence of McrC. Since ~18 kb has been deleted in the host strain used in Table 1, this putative gene must be some distance away. We have no evidence for or against the existence of a second specificity gene.

Despite the similarities, there are significant differences between the type I *hsdRMS* system and *mcrBC*. The lack of a modification subunit is an obvious one. More significantly, the *hsdRMS* and *mcrBC* systems are not detectably similar in polypeptide sequence, and cofactor requirements are similar but different, since we have preliminary evidence that McrBC activity in vitro requires GTP absolutely, rather than ATP (E. Sutherland and E. Raleigh, unpublished results). The nucleotide-binding site motif in McrB is thus of particular interest. It is found in a region of the protein where the sequence is in conflict with the data of others (53). We have shown our sequence to be correct (Fig. 4). For two reasons, we believe that the motif is a conserved, functionally significant region of the protein. First, as mentioned above, we have in vitro evidence for GTP dependence, to be presented elsewhere. Second, we searched the NBRF data base for proteins with greater sequence similarity to the McrB GXXXXGK (domain I) sequence than the three invariant amino acids and asked what the functional properties of that set might be. We found 10 proteins with matches at six of seven positions to the specific McrB sequence (GPPGVGK). All but one had functions plausibly related to nucleotide binding. Six are kinases, and three are likely to act in nucleic acid metabolism (the product of the *ruvB* DNA repair gene of *E. coli*, the phage Mu B protein, and a rhinovirus polyprotein). The 10th, porcine valosin, is of unknown function. A literature search turned up an 11th match: the yeast CDC6 protein, which is a proposed GTP-binding protein (30). This increases the likelihood that the sequence we have identified represents a functional nucleotide binding motif and not a random match.

There are other models equally consistent with the data, but all must take account of the ability of McrB to mediate restriction in the absence of McrC. For example, if the analogy with type I/type III systems is less strict, it is possible that McrB both recognizes and acts on a sequence present in λ · *Msp*I and that McrC cooperates with McrB to relax its sequence specificity and thus allows it to act on sequences present in λ · *Hae*II, T4*gt*, and the rest of the targets shown in Table 2. It is also possible that the analogy is altogether wrong and that McrB and McrC act sequentially in some way. For example, McrC might act by altering the target DNA in some way to make it susceptible to McrB action. We do not favor models in which McrB itself acts as a specificity subunit in association with a restriction function from elsewhere.

Technical complexities and data conflicts. Both the vector and the host strain used for genetic analysis can confound interpretation. We believe that these two factors contributed to the contradiction between the picture we derive here and the interpretations of Ross et al. (52), who concluded that McrB restricts ^{hm}C-containing DNA and that McrBC restricts ^{m5}C-containing DNA. These workers based their interpretations on work done with the high-copy-number vector pUC8, the ancestor of pEMBL19m, and with WA802 (also known as K802), the original *mcrB1* strain. We find both of these factors to be problematic.

Genetic constitution of *mcrB1* host strains. Ross et al. (52) observed low-level (10-fold) restriction of T4*gt* in an *mcrB1* host complemented by plasmids carrying *mcrB* alone. From this finding they concluded that McrB recognizes ^{hm}C and is responsible by itself for restriction of T4*gt*. We also found that plasmids carrying *mcrB* alone partially restore restriction of T4*gt* to an *mcrB1* strain. However, since these plasmids have no effect in a deletion strain (Table 1), we conclude that such strains express McrC at a low level, that McrC is limiting for restriction, and that partial complementation between plasmid-encoded McrB and chromosomally encoded McrC leads to restoration of low-level restriction. Ross et al. (52) were unable to distinguish the low-level restriction of *Alu*I-methylated pACYC in this situation from lack of restriction, just as we obtained unimpressive (but reproducible) restriction of λ · *Hae*II (Table 1). A consistent model is that the *mcrB1* mutation is in *mcrB*, inactivates the McrB protein, and reduces expression of the McrC protein, as a nonsense or frameshift mutation might do via polarity (63). This possibility is consistent with the result (M. Noyer-Weidner, personal communication) that synthesis of the ~49- and ~29-kDa McrB proteins is abolished and that of the 38- to 40-kDa McrC protein is reduced in in vitro transcription-translation assays of clones carrying this allele.

Vector artifact. Use of a single, high-copy-number vector for complementation experiments is also problematic. We find that a cloned fragment expressing very nearly wild type activity in one context expresses almost none in another (Table 3). We had difficulty characterizing constructs in the

very-high-copy-number vector, pEMBL19m, and finally abandoned the attempt, preferring to transfer constructs to other (low-copy-number) vectors in which phenotypes conferred could be interpreted unambiguously. We do not have a good explanation for the vector effect but view with caution experiments conducted in only one vector background. The experiments of Ross et al. (52) are also difficult to interpret because of the use of two incompatible, high-copy-number plasmids carrying different drug resistances in the same cell for complementation experiments. Relative copy levels are impossible to assess in such a situation.

Sequence disagreements. A major portion of the sequence of McrB reported here, encompassing the first element of the GTP-binding motif (Fig. 3 and text), conflicts with the sequence reported by Ross et al. (53). We used restriction digests (Fig. 4) to verify our sequence at 11 of the 15 conflicting positions. We also sequenced this region from a clone derived from the same strain as was their clone and obtained the same sequence as from our own clone. We believe that this result rules out natural sequence variation from strain to strain as a source of the conflict. A potential alternating DNA secondary structure (dyads B and C; Fig. 3) identified within the region of conflict may explain difficulty in reading the sequence of this region.

Clues to the evolutionary history of the immigration control region. G+C-rich potential transcription terminators (dyads A, D, E, and F in Fig. 3) flank an *mcrBC* sequence abnormally low in G+C content (40%; Fig. 5) compared with average *E. coli* sequences (50%), both coding and noncoding. The striking pattern of G+C variation suggests to us that this sequence has been recruited from foreign sources recently in the evolutionary history of the chromosome. It is also intriguing that the local low points in the G+C content of *hsdS* coincide with the parts of the protein known to be nonconserved (20), as though importation of foreign sequences were used to generate diversity. If the G+C content does reflect recent acquisition, the higher G+C content of the 700-bp intergenic region between *hsdS* and *mcrB* in K-12 would reflect fixation of that segment in the *E. coli* genome at relatively ancient times. This view is consistent with Southern blot hybridization data showing that the intergenic region is conserved in *E. coli* 15 T^-, whereas the flanking restriction loci (*hsd* and *mcr*) are not (12).

Particularly intriguing in view of this fluidity of sequence arrangement is the similarity between two of the dyad elements (A and E) just at the boundaries of the low-G+C segment, as though these represented a marker of sequence acquisition. An additional similar element is found between *hsdR* and *hsdM* (Figs. 5 and 6), although its position here is less suggestive.

We find attractive the idea that the *mcrBC* sequence represents a mobile element, but the sequence data do not provide unequivocal support for it. Mobile accessory elements frequently protect themselves from outside transcription with terminators near the ends of the element (13), as we see here. However, unlike many mobile elements of bacteria and higher organisms (reviewed in reference 4), the *mcrBC* sequence features no long (8 to 40 bases) terminal inverted repeat or associated short (2 to 13 bases) direct repeat. Nevertheless, some integrated mobile elements, including some transposons found in gram-positive bacteria but capable of insertion in *E. coli* (39), do not exhibit this arrangement, so the possibility remains open.

ACKNOWLEDGMENTS

We thank Mario Noyer-Weidner for useful discussions and for sharing information before publication, Geoff Wilson and Joe Heitman for plasmids and discussions, and Pamela Briggs for assistance with the sequence verification digests. Marjorie Russel gave valued advice in setting up the f1 infection assay. Helen Revel has been a continuing source of helpful criticism and encouragement. We thank Bob Blumenthal for critical discussions of the vector effect and Joan Brooks, Nancy Kleckner, and Ira Schildkraut for critical review of the manuscript, which made it much better.

LITERATURE CITED

1. **Abel, T., and T. Maniatis.** 1989. Action of leucine zippers. Nature (London) **341**:24–25.
2. **Ansorge, W., and S. Labeit.** 1984. Field gradients improve resolution on DNA sequencing gels. J. Biochem. Biophys. Methods **10**:237–243.
3. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. Greene Publishing Associates and John Wiley & Sons, Inc., New York.
4. **Berg, D. E., and M. M. Howe (ed.).** 1989. Mobile DNA. American Society for Microbiology, Washington, D.C.
5. **Berg, J. M.** 1986. Potential metal-binding domains in nucleic acid binding proteins. Science **232**:485–487.
6. **Bickle, T. A.** 1982. The ATP-dependent restriction endonucleases, p. 85–108. *In* S. M. Linn and R. J. Roberts (ed.), Nucleases. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
7. **Blumenthal, R. M., S. A. Gregory, and J. S. Cooperider.** 1985. Cloning of a restriction-modification system from *Proteus vulgaris* and its use in analyzing a methylase-sensitive phenotype in *Escherichia coli*. J. Bacteriol. **164**:501–509.
8. **Bolivar, F., R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heynecker, H. W. Boyer, J. H. Crosa, and S. Falkow.** 1977. Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95–113.
9. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41**:459–472.
10. **Butkus, V., S. Klimasauskas, Z. Petrauskiene, Z. Maneliene, A. Lebionka, and A. Janulaitis.** 1987. Interaction of *Alu*I, *Cfr*6I, and *Pvu*II restriction-modification enzymes with substrates containing either N^4-methylcytosine or 5-methylcytosine. Biochim. Biophys. Acta **909**:201–207.
11. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134**: 1141–1156.
12. **Daniel, A. S., F. V. Fuller-Pace, D. M. Legge, and N. E. Murray.** 1988. Distribution and diversity of *hsd* genes in *Escherichia coli* and other enteric bacteria. J. Bacteriol. **170**:1775–1782.
13. **Davis, M. A., R. W. Simons, and N. Kleckner.** 1985. Tn*10* protects itself at two levels from fortuitous activation by external promoters. Cell **43**:379–387.
14. **Davis, R. W., D. Botstein, and J. R. Roth.** 1980. Advanced bacterial genetics: a manual for genetic engineering. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Dente, L., G. Cesareni, and R. Cortese.** 1983. pEMBL: a new family of single-stranded plasmids. Nucleic Acids Res. **11**:1645–1655.
16. **Dever, T. E., M. J. Glynias, and W. C. Merrick.** 1987. GTP-binding domain: three consensus sequence elements with distinct spacing. Proc. Natl. Acad. Sci. USA **84**:1814–1818.
17. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
18. **Dila, D., and E. A. Raleigh.** 1988. Genetic dissection of the methylcytosine-specific restriction system *mcrB* of *Escherichia coli* K-12. Gene **74**:23–24.
19. **Georgopoulos, K.** 1967. Isolation and preliminary characterization of T4 mutants with non-glucosylated DNA. Biochem. Biophys. Res. Commun. **28**:179–184.

20. **Gough, J. A., and N. E. Murray.** 1983. Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. **166**:1–19.
21. **Haltiner, M., T. Kempe, and R. Tjian.** 1985. A novel strategy for constructing clustered point mutations. Nucleic Acids Res. **13**:1015–1025.

21a. **Hattori, M., and Y. Sakaki.** 1986. Dideoxy sequencing method using denatured plasmid templates. Anal. Biochem. **152**:232–238.

22. **Heitman, J., and P. Model.** 1987. Site-specific methylases induce the SOS DNA repair response in *Escherichia coli*. J. Bacteriol. **169**:3243–3250.
23. **Heitman, J., J. Treisman, N. G. Davis, and M. Russel.** 1989. Cassettes of the f1 intergenic region. Nucleic Acids Res. **17**:4413.
24. **Howard, K. A., C. Card, J. S. Benner, H. L. Callahan, R. Maunus, K. Silber, G. Wilson, and J. E. Brooks.** 1986. Cloning the *Dde*I restriction-modification system using a two-step method. Nucleic Acids Res. **14**:7939–7951.
25. **Keller, C., N. Corcoran, and R. J. Roberts.** 1984. Computer programs for handling nucleic acid sequences. Nucleic Acids Res. **12**:379–386.
26. **Kiss, A., F. Posfai, C. C. Keller, P. Venetianer, and R. J. Roberts.** 1985. Nucleotide sequence of the *Bsu*RI restriction-modification system. Nucleic Acids Res. **13**:6403–6421.
27. **Kouzarides, T., and E. Ziff.** 1989. Leucine zippers of *fos*, *jun* and GCN4 dictate dimerization specificity and thereby control DNA binding. Nature (London) **340**:568–571.
28. **Lacks, S. A., B. M. Mannarelli, S. S. Springhorn, and B. Greenberg.** 1986. Genetic basis of the complementary *Dpn*I and *Dpn*II restriction systems of *S. pneumoniae*: an intercellular cassette mechanism. Cell **46**:993–1000.
29. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759–1764.
30. **Lisziewicz, J., A. Godany, D. V. Agoston, and H. Küntzel.** 1988. Cloning and characterization of the *Saccharomyces cerevisiae* CDC6 gene. Nucleic Acids Res. **16**:11507–11520.
31. **Loenen, W. A. M., A. S. Daniel, H. D. Braymer, and N. E. Murray.** 1987. Organization and sequence of the *hsd* genes of *Escherichia coli* K-12. J. Mol. Biol. **198**:159–170.
32. **Lupski, J. R., S. J. Projan, L. S. Ozaki, and G. N. Godson.** 1986. A temperature-dependent pBR322 copy number mutant resulting from a Tn5 position effect. Proc. Natl. Acad. Sci. USA **83**:7381–7385.
33. **Luria, S. E., and M. L. Human.** 1952. A nonhereditary, host-induced variation of bacterial viruses. J. Bacteriol. **64**:557–559.
34. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Messing, J., B. Gronenborn, B. Müller-Hill, and P. H. Hofschneider.** 1977. Filamentous phage M13 as a cloning vehicle: insertion of a *Hin*dII fragment of the *lac* regulatory region in M13 replicative form in vitro. Proc. Natl. Acad. Sci. USA **74**:3642–3646.
36. **Miki, T., T. Yasukochi, H. Nagatani, M. Furuno, T. Orita, H. Yamada, T. Imoto, and T. Horiuchi.** 1987. Construction of a plasmid vector for the regulatable high level expression of eukaryotic genes in *Escherichia coli*: an application to overproduction of chicken lysozyme. Protein Eng. **1**:327–332.
37. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. EMBO J. **4**:1609–1614.
38. **Modrich, P., and R. J. Roberts.** 1982. Type II restriction and modification enzymes, p. 109–154. *In* S. Linn and R. J. Roberts (ed.), Nucleases. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **Murphy, E.** 1989. Transposable elements in gram-positive bacteria, p. 269–288. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
40. **Noyer-Weidner, M., R. Diaz, and L. Reiners.** 1986. Cytosine-specific DNA modification interferes with plasmid establishment in *Escherichia coli* K-12: involvement of *rglB*. Mol. Gen. Genet. **205**:469–475.
41. **Nwankwo, D. O., and G. G. Wilson.** 1988. Cloning and expression of the *Msp*I restriction and modification genes. Gene **64**:1–8.
42. **Olsson, A., T. Moks, M. Uhlen, and A. B. Gaal.** 1984. Uniformly spaced banding pattern in DNA sequencing gels by use of field-strength gradient. J. Biochem. Biophys. Methods **10**:83–90.
43. **O'Shea, E. K., R. Rutkowski, and P. S. Kim.** 1989. Evidence that the leucine zipper is a coiled coil. Science **243**:538–542.
44. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
45. **Petrusyte, M., J. Bitinaite, S. Menkevicius, S. Klimasauskas, V. Butkus, and A. Janulaitis.** 1988. Restriction endonucleases of a new type. Gene **74**:89–91.
46. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
47. **Raleigh, E. A., N. E. Murray, H. Revel, R. M. Blumenthal, D. Westaway, A. D. Reith, P. W. J. Rigby, J. Elhai, and D. Hanahan.** 1988. McrA and McrB restriction phenotypes of some *E. coli* strains and implications for gene cloning. Nucleic Acids Res. **16**:1563.
48. **Raleigh, E. A., R. Trimarchi, and H. Revel.** 1989. Genetic and physical mapping of the *mcrA* (*rglA*) and *mcrB* (*rglB*) loci of *Escherichia coli* K-12. Genetics **122**:279–296.
49. **Raleigh, E. A., and G. Wilson.** 1986. *Escherichia coli* K-12 restricts DNA containing 5-methylcytosine. Proc. Natl. Acad. Sci. USA **83**:9070–9074.
50. **Revel, H. R.** 1983. DNA modification: glucosylation, p. 156–165. *In* C. K. Mathews, E. M. Kutter, G. Mosig, and P. Berget (ed.), Bacteriophage T4. American Society for Microbiology, Washington, D.C.
51. **Ross, T. D., E. C. Achberger, and H. D. Braymer.** 1987. Characterization of the *Escherichia coli* modified cytosine restriction (*mcrB*) gene. Gene **61**:277–289.
52. **Ross, T. K., E. C. Achberger, and H. D. Braymer.** 1989. Identification of a second polypeptide required for McrB restriction of 5-methylcytosine-containing DNA in *Escherichia coli* K-12. Mol. Gen. Genet. **216**:402–407.
53. **Ross, T. K., E. C. Achberger, and H. D. Braymer.** 1989. Nucleotide sequence of the McrB region of *Escherichia coli* K-12 and evidence for two independent translational initiation sites at the *mcrB* locus. J. Bacteriol. **171**:1974–1981.
54. **Russel, M., S. Kidd, and M. R. Kelley.** 1986. An improved filamentous helper phage for generating single-stranded plasmid DNA. Gene **45**:333–338.
55. **Sain, B., and N. E. Murray.** 1980. The *hsd* (host specificity) genes of *E. coli* K-12. Mol. Gen. Genet. **180**:35–46.
56. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
57. **Schumperli, D., K. McKenney, D. A. Sobieski, and M. Rosenberg.** 1982. Translational coupling at an intercistronic boundary of the *Escherichia coli* galactose operon. Cell **30**:865–871.
58. **Shapira, S. K., J. Cou, F. V. Richaud, and M. J. Casadaban.** 1983. New versatile plasmid vectors for expression of hybrid proteins coded by a cloned gene fused to *lacZ* gene sequences encoding an enzymatically active carboxy-terminal portion of β-galactosidase. Gene **25**:71–82.
59. **Smith, H. O., and S. V. Kelley.** 1984. Methylases of type II restriction-modification systems, p. 39–71. *In* A. Razin, H. Cedar, and A. D. Riggs (ed.), DNA methylation: biochemistry and biological significance. Springer-Verlag, New York.
60. **Tinoco, I., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenback, D. M. Crothers, and J. Gralla.** 1973. Improved estimation of secondary structure in ribonucleic acids. Nature (London) New Biol. **246**:40–41.
61. **Williams, S. A., B. E. Slatko, L. S. Moran, and S. M. DeSimone.** 1986. Sequencing in the fast lane: a rapid protocol for [α-^{35}S] dATP dideoxy DNA sequencing. BioTechniques **4**:138–147.
62. **Zinder, N. D., and J. D. Boeke.** 1982. The filamentous phage (Ff) as vectors for recombinant DNA—a review. Gene **19**:1–10.
63. **Zipser, D.** 1970. Polarity and translational punctuation, p. 221–232. *In* J. R. Beckwith and D. Zipser (ed.), The lactose operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.